



TITLE:

Development of correlation-based clustering method and its application to software sensing

AUTHOR(S):

Fujiwara, Koichi; Kano, Manabu; Hasebe, Shinji

CITATION:

Fujiwara, Koichi ...[et al]. Development of correlation-based clustering method and its application to software sensing. Chemometrics and Intelligent Laboratory Systems 2010, 101(2): 130-138

ISSUE DATE:

2010-04-15

URL:

<http://hdl.handle.net/2433/120792>

RIGHT:

© 2010 Elsevier B.V.; This is not the published version. Please cite only the published version.; この論文は出版社版ではありません。引用の際には出版社版をご確認ご利用ください。

Development of Correlation-based Clustering Method and Its Application to Software Sensing

Koichi Fujiwara*, Manabu Kano, Shinji Hasebe*

Department of Chemical Engineering, Kyoto University, Nishikyo-ku, Kyoto 615-8510, Japan

Abstract

The individuality of production devices should be taken into account when soft-sensors are designed for parallelized devices. Since it is expressed as differences of the correlation among measured variables, it is useful to cluster samples on the basis of the correlation among variables for adopting a multi-model approach. In addition, changes in process characteristics can be coped with in the same way. In the present work, a new clustering method, referred to as NC-spectral clustering, is proposed by integrating the nearest correlation (NC) method and spectral clustering. Spectral clustering is a graph partitioning method that can be used for sample classification when an affinity matrix of a weighted graph is given. The NC method can detect samples that are similar to the query from the viewpoint of the correlation without a teacher signal. In the proposed method, the NC method is used for constructing the weighted graph that expresses the correlation-based similarities between samples and the constructed graph is partitioned by using spectral clustering. In addition, a new soft-sensor design method is proposed on the basis of the proposed NC-spectral clustering. The usefulness of the proposed methods is demonstrated through a numerical example and a case study of parallelized batch processes. The performance of the proposed correlation-based method is better than that of the conventional distance-based methods.

Key words: Spectral clustering, Nearest correlation method, Soft-sensor, Batch process, Correlation, Graph Theory

1. Introduction

Soft-sensors have been widely used for estimating product quality or other key variables. Partial least squares (PLS) and artificial neural network (ANN) have been widely accepted as useful techniques for soft-sensor design [1–5]. In addition, the application of subspace identification (SSID) to soft-sensor design has been reported [6, 7].

Even if a good soft-sensor is developed successfully, its estimation performance deteriorates as process characteristics change. In chemical processes, process characteristics are changed by catalyst deactivation, scale adhesion and so on. In semiconductor manufacturing processes, periodic cleaning of equipment changes the process characteristics dramatically. Therefore, maintenance of soft-sensors is very important to keep their estimation performance. They should be updated as the

process characteristics change. However, repeated and manual construction of them should be avoided due to heavy workload [8].

To cope with changes in process characteristics and to update statistical models automatically, recursive PLS [9] and Just-In-Time (JIT) modeling [10–12] have been proposed. Recursive PLS updates a PLS model recursively, and JIT modeling builds a local model from neighbor samples around the query only when an estimate is requested. Although these methods can adapt models to new operating conditions, they cannot always achieve high estimation performance.

Recently, a new JIT modeling method, referred to as Correlation-based JIT (CoJIT) modeling, has been proposed [13]. Since changes in process characteristics are expressed as differences of the correlation among variables, CoJIT modeling builds a local model from samples whose correlation can properly describe the query.

In addition, the individuality of production devices should be taken into account. In semiconductor processes, for example, tens of parallelized production devices are used, and they have different characteristics

*Corresponding author. Tel.: +81-75-383-2677; fax: +81-75-383-2657.

Email address: fujiwara@cheme.kyoto-u.ac.jp (Koichi Fujiwara)

Preprint submitted to Chemometrics and Intelligent Laboratory Systems

February 19, 2010

even if their catalog specifications are the same. In such a case, a soft-sensor developed for one device is not always applicable to another device, and it is very laborious to construct soft-sensors for each device. Therefore, a practical soft-sensor design method that can cope with the individuality of production devices should be developed.

If there are some devices whose characteristics are similar to each other, the same soft-sensor may be applicable to them. That is, to construct soft-sensors that can cope with the individuality of production devices, it is useful to cluster operation data of parallelized processes into a few classes according to their characteristics and to construct a model for each class. In addition, the number of developed soft-sensors can be reduced by comparison with the case where soft-sensors are constructed for each device.

The k -means method has been used for sample classification. Although it can cluster samples on the basis of the distance, it does not take into account the correlation among variables. Recently, self-organizing map (SOM) has been proposed [14]. SOM is a machine learning process that imitates the brain learning process, and it can visualize high dimensional data as a map on the basis of similarities between samples. However, SOM does not always give clear boundaries between clusters on the map. In addition, it requires high computational load, and its parameter tuning and data preprocessing are complicated.

In the present work, a new clustering method, referred to as NC-spectral clustering, is proposed. In the proposed method, the nearest correlation (NC) method [15] that can detect samples whose correlation is similar to the query and spectral clustering [16, 17] that can partition a weighted graph are integrated. The proposed NC-spectral clustering can classify samples according to their correlation among variables without teacher signals. In addition, a new soft-sensor design method based on NC-spectral clustering is developed. The usefulness of the proposed methods is demonstrated through a case study of parallelized batch processes.

2. Spectral clustering

Spectral clustering is a clustering method on the basis of the graph theory. It can partition a weighted graph, whose weights express affinities between nodes, into subgraphs through cutting some of their arcs.

Although some spectral clustering algorithms have been proposed, the Max-Min Cut (Mcut) method [16] is described in this section.

Given a weighted graph G and its adjacency matrix (affinity matrix) W , G is partitioned into two subgraphs A and B . The affinity between A and B is defined as

$$\text{cut}(A, B) \equiv W(A, B) \quad (1)$$

$$W(A, B) = \sum_{u \in A, v \in B} W_{u,v} \quad (2)$$

$$W(A) \equiv W(A, A). \quad (3)$$

where u and v denote nodes of subgraphs A and B , respectively. That is, the affinity between subgraphs $\text{cut}(A, B)$ is the sum of the weights of the arcs between subgraphs. The Mcut method searches subgraphs A and B that can minimize $\text{cut}(A, B)$ and maximize $W(A)$ and $W(B)$, simultaneously. The objective function of the Mcut method is as follows:

$$\min J = \frac{\text{cut}(A, B)}{W(A)} + \frac{\text{cut}(A, B)}{W(B)}. \quad (4)$$

Since indexes of nodes are interchangeable, the affinity matrix W can be defined as follows:

$$W = \begin{bmatrix} W_A & W_{A,B} \\ W_{B,A} & W_B \end{bmatrix} \quad (5)$$

where W_A and W_B are the adjacency matrices within the subgraphs A and B , respectively, and $W_{A,B}$ and $W_{B,A}$ are the affinity matrix between A and B . In addition, the vectors $\mathbf{x} = [1, \dots, 1, 0, \dots, 0]^T$ and $\mathbf{y} = [0, \dots, 0, 1, \dots, 1]^T$ that express partition into subgraphs are satisfied as follows:

$$W(A, B) = \mathbf{x}^T (\mathbf{D} - \mathbf{W}) \mathbf{y} = \mathbf{y}^T (\mathbf{D} - \mathbf{W}) \mathbf{x} \quad (6)$$

$$W(A) = \mathbf{x}^T \mathbf{W} \mathbf{x} \quad (7)$$

$$W(B) = \mathbf{y}^T \mathbf{W} \mathbf{y} \quad (8)$$

where $\mathbf{D} = \text{diag}(\mathbf{W}\mathbf{e})$ and $\mathbf{e} = [1, \dots, 1]^T$. Using these equations, Eq. (4) can be rewritten as

$$\min_{\mathbf{x}, \mathbf{y}} J = \frac{\mathbf{x}^T (\mathbf{D} - \mathbf{W}) \mathbf{x}}{\mathbf{x}^T \mathbf{W} \mathbf{x}} + \frac{\mathbf{y}^T (\mathbf{D} - \mathbf{W}) \mathbf{y}}{\mathbf{y}^T \mathbf{W} \mathbf{y}}. \quad (9)$$

Only the first term of Eq. (9) needs to be analyzed.

The minimization problem of Eq. (9) is hard to solve, since the optimizing variables \mathbf{x} and \mathbf{y} are binary variables. Therefore, these binary variables should be relaxed. The new index vector $\mathbf{q} = \{a, -b\}$, ($a, b > 0$) is introduced as follows:

$$q_u = \begin{cases} a, & (u \in A) \\ -b, & (u \in B) \end{cases}. \quad (10)$$

Finally, the problem can be written as follows:

$$\min_{\mathbf{q}} J_q = \frac{\mathbf{q}^T (\mathbf{D} - \mathbf{W}) \mathbf{q}}{\mathbf{q}^T \mathbf{W} \mathbf{q}}. \quad (11)$$

This minimization problem results in the eigenvalue problem.

$$(\mathbf{I} - \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}) \mathbf{z} = \lambda \mathbf{z} \quad (12)$$

The solution \mathbf{q}^* is expressed as $\mathbf{q}^* = \mathbf{D}^{-1/2} \mathbf{z}_2$, where \mathbf{z}_2 is the eigenvector corresponding to the second largest eigenvalue λ_2 ; λ_2 and \mathbf{z}_2 are called Fiedler value and Fiedler vector, respectively [18]. Through the above procedure, the Mcut method does not need parameter tuning and its calculation is fast.

Since the Mcut method partitions a weighted graph into two subgraphs, the procedure described above is repeated when three or more subgraphs are needed.

In spectral clustering, the definition of an affinity is arbitrary and affects results. Ng *et al.* defined the affinity between samples s_i and s_j by using the Gaussian kernel [17].

$$(W)_{i,j} = \exp\left(\frac{-d^2(s_i, s_j)}{2\sigma^2}\right) \quad (13)$$

where $d(\cdot, \cdot)$ is a distance function and σ is a tuning parameter.

3. NC-spectral clustering

In the present work, a new clustering method based on the correlation among variables is proposed. In the proposed method, the correlation-based affinities between samples are calculated by using the nearest correlation (NC) method to construct a weighted graph, and the constructed weighted graph is partitioned by spectral clustering. This method is referred to as NC-spectral clustering.

3.1. Nearest correlation method

The NC method can detect samples whose correlation is similar to the query without any teacher signals [15].

The concept of the NC method is as follows: Suppose that the affine subspace P in Fig. 1 (left) shows the correlation among variables and all the samples on P have the same correlation. Although $\mathbf{x}_1, \dots, \mathbf{x}_5$ have the same correlation, \mathbf{x}_6 and \mathbf{x}_7 have a different correlation from the others. The NC method aims to detect samples whose correlation is similar to the query \mathbf{x}_1 . In this example, $\mathbf{x}_2, \dots, \mathbf{x}_5$ on P should be detected.

At first, the whole space is translated so that the query becomes the origin as shown in Fig. 1 (right). That is, \mathbf{x}_1 is subtracted from all other samples \mathbf{x}_i ($i = 2, \dots, 7$). Since the translated affine subspace contains the origin, it becomes the linear subspace V .

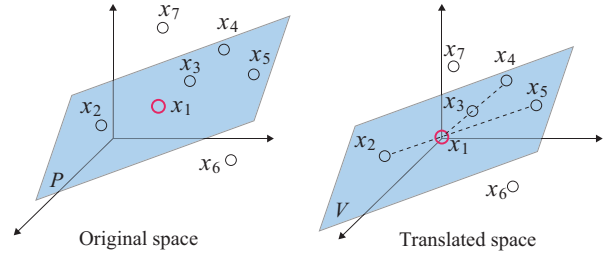


Figure 1: An example of the procedure of the NC method

Next, a line connecting each sample and the origin is drawn. Suppose another sample can be found on this line. In this case, $\mathbf{x}_2 - \mathbf{x}_5$ and $\mathbf{x}_3 - \mathbf{x}_4$ satisfy such a relationship as shown in Fig. 1 (right). The correlation coefficients of these pairs must be 1 or -1 . On the other hand, \mathbf{x}_6 and \mathbf{x}_7 that are not the elements of V cannot make such pairs. Therefore, the pairs whose correlation coefficients are ± 1 are thought to have the same correlation as \mathbf{x}_1 .

In practice, the threshold of the correlation coefficient γ ($0 < \gamma \leq 1$) has to be used, since there are no pairs whose correlation coefficient is strictly ± 1 . Therefore, the pairs should be selected when the absolute values of their correlation coefficients are larger than γ .

Using the above procedure, the pairs whose correlation is similar to the query can be detected.

3.2. NC-spectral clustering

The correlation-based affinity matrix for spectral clustering can be constructed by using the NC method. Assume that samples $\mathbf{x}_n \in \mathbb{R}^M$ ($n = 1, \dots, N$) are stored in the database. The procedure of the proposed NC-spectral clustering is as follows:

- Step 1. Set the zero matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$, γ ($0 < \gamma \leq 1$) and $L = 1$.
- Step 2. Set the zero matrix $\mathbf{S}_L \in \mathbb{R}^{N \times N}$.
- Step 3. $\mathbf{x}'_n = \mathbf{x}_n - \mathbf{x}_L$ for $n = 1, 2, \dots, N$ ($n \neq L$).
- Step 4. Calculate the correlation coefficients $C_{k,l}$ between all possible pairs of \mathbf{x}'_k and \mathbf{x}'_l ($k \neq l$).
- Step 5. Select all the pairs of \tilde{k} and \tilde{l} satisfying $|C_{\tilde{k},\tilde{l}}| \geq \gamma$.
- Step 6. $(\mathbf{S}_L)_{\tilde{k},\tilde{l}} = (\mathbf{S}_L)_{\tilde{l},\tilde{k}} = 1$.
- Step 7. $\mathbf{S} = \mathbf{S} + \mathbf{S}_L$.
- Step 8. If $L = N$, output \mathbf{S} as the affinity matrix. Otherwise, $L = L + 1$ and return to 2.
- Step 9. Partition the graph expressed by \mathbf{S} through spectral clustering.

In the above procedure, steps 3~5 correspond to the NC method.

3.3. Illustrative example

The detailed function of the proposed method is illustrated through a simple example. An objective data set consist of nine samples $\mathbf{x}_1, \dots, \mathbf{x}_9 \in \mathbb{R}^2$; $\mathbf{x}_1, \dots, \mathbf{x}_4$ and $\mathbf{x}_5, \dots, \mathbf{x}_8$ are on the lines l and k , respectively, as shown in Fig. 2. On the other hand, \mathbf{x}_9 is an outlier. That is, the data set consists of three classes $\{\mathbf{x}_1, \dots, \mathbf{x}_4\}$, $\{\mathbf{x}_5, \dots, \mathbf{x}_8\}$ and $\{\mathbf{x}_9\}$. In addition, samples \mathbf{x}_2 , \mathbf{x}_7 and \mathbf{x}_9 are arranged in line by chance.

First, the zero matrix $\mathbf{S} \in \mathbb{R}^{9 \times 9}$ is defined, and the correlations of all possible pairs of samples are checked by the NC method. For example, when \mathbf{x}_1 is the query, \mathbf{x}_2 - \mathbf{x}_3 , \mathbf{x}_2 - \mathbf{x}_4 and \mathbf{x}_3 - \mathbf{x}_4 are detected as pairs whose correlation is similar to \mathbf{x}_1 , and $(\mathbf{S})_{2,3} = (\mathbf{S})_{3,2} = 1$, $(\mathbf{S})_{2,4} = (\mathbf{S})_{4,2} = 1$, $(\mathbf{S})_{3,4} = (\mathbf{S})_{4,3} = 1$.

In the same way, since \mathbf{x}_1 - \mathbf{x}_3 , \mathbf{x}_1 - \mathbf{x}_4 , \mathbf{x}_3 - \mathbf{x}_4 and \mathbf{x}_7 - \mathbf{x}_9 are detected when \mathbf{x}_2 is the query, one is added to the elements of \mathbf{S} corresponding to these pairs. At this time, $(\mathbf{S})_{3,4} = (\mathbf{S})_{4,3} = 2$ because the pair \mathbf{x}_3 - \mathbf{x}_4 is detected again. This procedure is repeated so that all samples become the query. Finally, the affinity matrix \mathbf{S} becomes as follows:

$$\mathbf{S} = \begin{bmatrix} 0 & 2 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & 2 & 0 & 0 & 1 & 0 & 1 \\ 2 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 2 & 2 & 0 \\ 0 & 1 & 0 & 0 & 2 & 2 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 & 2 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}. \quad (14)$$

In \mathbf{S} , $(\mathbf{S})_{2,7} = (\mathbf{S})_{7,2} = 1$, $(\mathbf{S})_{2,9} = (\mathbf{S})_{9,2} = 1$ and $(\mathbf{S})_{7,9} = (\mathbf{S})_{9,7} = 1$ since samples \mathbf{x}_2 , \mathbf{x}_7 and \mathbf{x}_9 are arranged in line by chance and the pairs of these samples are detected by the NC method. However, the weights of these pairs in \mathbf{S} are smaller than those of the pairs that have the true correlation.

Figure 3 shows an example of the graph expression of the calculated affinity matrix \mathbf{S} . In Fig. 3, the length of each arc is inversely proportional to their weights. By partitioning this graph using spectral clustering, the nodes are classified into $\{\mathbf{x}_1, \dots, \mathbf{x}_4\}$, $\{\mathbf{x}_5, \dots, \mathbf{x}_8\}$ and $\{\mathbf{x}_9\}$. This example clearly shows that the proposed NC-spectral clustering can classify samples on the basis of the correlation among variables.

3.4. Numerical examples

The discrimination performance of the proposed NC-spectral clustering is compared with that of the k -means method and spectral clustering through numerical examples.

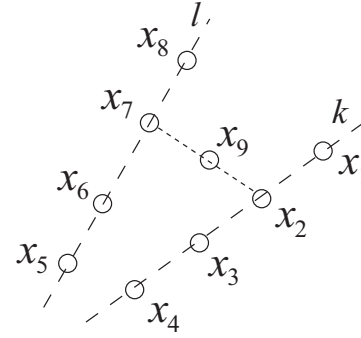


Figure 2: An objective data set

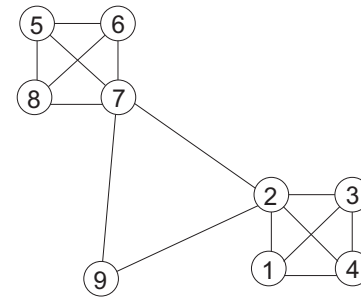


Figure 3: A weighted graph expressing the affinity matrix \mathbf{S}

3.4.1. Two-dimensional case

The discrimination performances are compared through two-dimensional examples. The objective data set consists of three classes that have different correlation and intersect at the origin. 100 samples in each of three classes are generated by the following equation.

$$\mathbf{x} = s\mathbf{a}_i + \mathbf{n} \quad (i = 1, 2, 3) \quad (15)$$

where $s \sim N(0, 1)$, measurement noise $\mathbf{n} = [n_1 \ n_2]^T$, $n_j \sim N(0, 0.01)$, and $N(m, \sigma)$ is a random number following the normal distribution whose mean is m and standard deviation is σ . The coefficient matrices $\mathbf{a}_i \in \mathbb{R}^2$ are $\mathbf{a}_1 = [1 \ 2]^T$, $\mathbf{a}_2 = [2 \ 2]^T$ and $\mathbf{a}_3 = [2 \ 1]^T$.

In the conventional spectral clustering, the affinities between samples are defined using the Gaussian kernel in Eq. (13), the Euclidean distance is used as a distance function $d(\cdot, \cdot)$ and $\sigma = 1$. In the proposed method, the parameter of the NC method is $\gamma = 0.999$. These parameters are determined by trial and error.

The generated samples and the clustering results of the k -means method, spectral clustering and the proposed NC-spectral clustering are shown in Fig. 4. The conventional distance-based methods cannot classify samples correctly. On the other hand, the proposed method can classify samples accurately in most regions except around the origin.

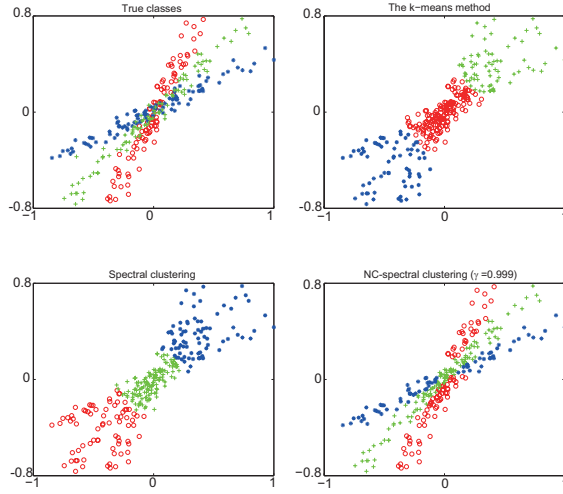


Figure 4: Classification results of the 2-dimensional example when three classes intersect at the origin: true classes (top-left), the k -means method (top-right), spectral clustering (bottom-left) and NC-spectral clustering (bottom-right)

To evaluate the effect of the parameter γ on the clustering results, samples are classified using NC-spectral clustering with different γ . Fig. 5 show the clustering results with $\gamma = 0.998, 0.995, 0.99$ and 0.98 . The clustering performance deteriorates as γ becomes small, and the clustering result does not change much when γ is less than 0.99 . This example shows the parameter γ should be close to one although the optimal value depends on the variance of measurement noise.

In addition, another case where three classes are located separately from each other and contaminated by large measurement noise is considered. 100 samples in each of three classes are generated by the following equation.

$$\mathbf{x} = s\mathbf{a}_i + \mathbf{b}_i + \mathbf{n}_i \quad (i = 1, 2, 3) \quad (16)$$

where s and \mathbf{a}_i are the same as Eq. (15). The biases $\mathbf{b}_i \in \mathbb{R}^2$ are $\mathbf{b}_1 = [0 \ 0.4]^T$, $\mathbf{b}_2 = [-0.5 \ 0]^T$ and $\mathbf{b}_3 = [0.5 \ 0.4]^T$, and the measurement noise is $\mathbf{n}_i = [n_1 \ n_2]^T$, $n_j \sim N(0, \sigma_j)$, where $\sigma_1 = 0.02$, $\sigma_2 = 0.03$ and $\sigma_3 = 0.04$.

The generated samples and the clustering results of the k -means method, spectral clustering and the proposed NC-spectral clustering are shown in Fig. 6. The settings of spectral clustering is the same as the previous example and the parameter of NC-spectral clustering is $\gamma = 0.999$.

In Fig. 6, every clustering method can discriminate samples accurately in most regions and the clustering results are similar to each other. This example shows

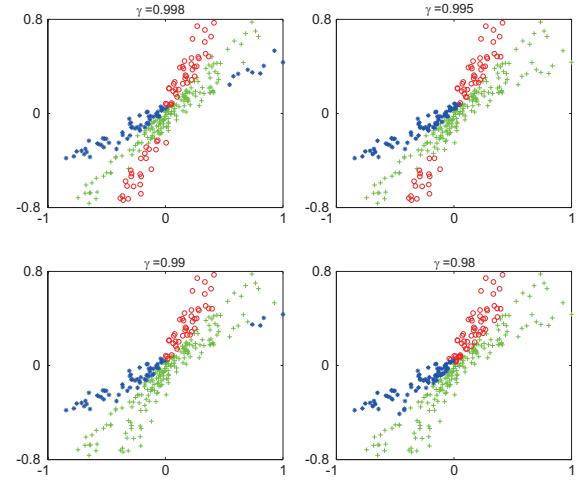


Figure 5: Classification results of NC-spectral clustering with different parameters γ : $\gamma = 0.998$ (top-left), 0.995 (top-right), 0.99 (bottom-left) and 0.98 (bottom-right)

that the proposed NC-spectral clustering also functions successfully as well as the conventional distance-based methods even when classes are located separately from each other.

3.4.2. Five-dimensional case

The discrimination performances are compared through a five-dimensional example. The objective data set consists of three classes that have different correlation and intersect at the origin. 100 samples in each of three classes are generated by the following equation.

$$\mathbf{x} = \mathbf{A}_i \mathbf{s} + \mathbf{n} \quad (i = 1, 2, 3) \quad (17)$$

where $\mathbf{s} = [s_1 \ s_2]$, $s_i \sim N(0, 1)$ and $\mathbf{n} = [n_1 \ n_2]^T$, $n_j \sim N(0, 0.01)$. The coefficient matrices $\mathbf{A}_i \in \mathbb{R}^{5 \times 2}$ are as follows:

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 1 \\ 2 & 3 \\ 1 & 3 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 3 & 3 \\ 2 & 1 \\ 3 & 1 \\ 3 & 2 \\ 2 & 0 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 1 & 3 \\ 0 & 4 \\ 3 & 1 \end{bmatrix}.$$

The discrimination rate is defined as

$$\text{Discrimination rate}[\%] = \frac{L}{K} \times 100 \quad (18)$$

where K is the number of detected samples and $K = 100$ in this example. L ($L \leq K$) is the number of samples that belong to the true class, out of K detected samples. The settings of spectral clustering is the same as Sec. 3.4.1 and the parameter of NC-spectral clustering is $\gamma = 0.999$ and 0.99 .

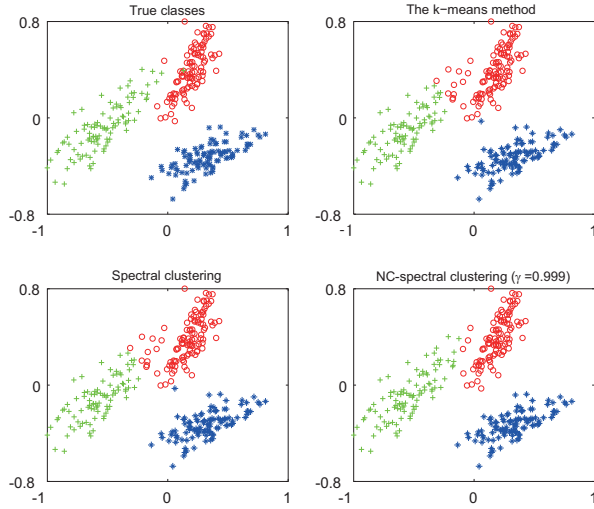


Figure 6: Classification results of the 2-dimensional example when three classes are located separately from each other: true classes (top-left), the k -means method (top-right), spectral clustering (bottom-left) and NC-spectral clustering (bottom-right)

Table 1 shows the discrimination performances of the k -means method, spectral clustering and NC-spectral clustering. The proposed NC-spectral clustering with $\gamma = 0.999$ can achieve higher discrimination performance than the other methods although the discrimination performance deteriorates in the case of $\gamma = 0.99$. These results clearly show that the proposed NC-spectral clustering can discriminate the correlation among variables when the parameter γ is approximately selected.

4. Soft-sensor design based on NC-spectral clustering

The operation data need to be clustered according to their characteristics to construct soft-sensors that can cope with the individuality of production devices. Since the individuality of production devices affects differences of the correlation among variables, the proposed

Table 1: Discrimination performances of the 5-dimensional example

	Discrimination rate [%]		
	Class 1	Class 2	Class 3
The k -means method	43	19	75
Spectral clustering	40	35	31
NC-spectral clustering ($\gamma = 0.999$)	85	84	93
NC-spectral clustering ($\gamma = 0.99$)	45	51	85

NC-spectral clustering can cluster the operation data according to their characteristics.

In the present work, a new soft-sensor design method based on NC-spectral clustering is proposed. In the proposed method, the operation data are clustered by using NC-spectral clustering and models are constructed for each class.

Assume that the input samples and output samples are stored in the database $\mathbf{x}_n \in \mathbb{R}^M$ and $\mathbf{y}_n \in \mathbb{R}^L$ ($n = 1, 2, \dots, N$). The proposed soft-sensor design procedure is as follows:

1. Classify the input samples \mathbf{x}_n to P classes using NC-spectral clustering, and $\Omega_j = \{n \mid \mathbf{x}_n \in K_j\}$, ($j = 1, 2, \dots, P$).
2. Construct models $f_j : X \rightarrow Y$ from \mathbf{x}_i and \mathbf{y}_i ($i \in \Omega_j$) for K_j , where X is the set of input and Y is that of output.
3. Classify a new sample $\tilde{\mathbf{x}}$ to class $K_j = h(\tilde{\mathbf{x}})$ when $\tilde{\mathbf{x}}$ is measured, where $h : X \rightarrow K$ is a classifier and K is the set of class.
4. Calculate an estimate $\hat{\mathbf{y}} = f_j(\tilde{\mathbf{x}})$.

In the above algorithm, any modeling method can be used for building a local model f . In the present work, PLS is used to cope with the collinearity problem.

The Q statistic can be used as the evaluation function of a classifier h . The Q statistic is derived from principal component analysis (PCA), which is a tool for data compression and information extraction [19]. PCA finds linear combinations of variables that describe major trends in a data set.

In PCA, the loading matrix $\mathbf{V}_R \in \mathbb{R}^{M \times R}$ is derived as the right singular matrix of a data matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$, and the column space of \mathbf{V}_R is the subspace spanned by principal components. Here, M , N , and $R(\leq M)$ denote the numbers of variables, samples, and principal components retained in the PCA model, respectively. All variables are mean-centered and appropriately scaled. The score is a projection of a sample $\mathbf{x} \in \mathbb{R}^M$ onto the subspace spanned by principal components. The score $\mathbf{t}_R \in \mathbb{R}^R$ is given by

$$\mathbf{t}_R = \mathbf{V}_R^T \mathbf{x}. \quad (19)$$

\mathbf{x} can be reconstructed or estimated from \mathbf{t}_R with linear transformation \mathbf{V}_R .

$$\hat{\mathbf{x}} = \mathbf{V}_R \mathbf{t}_R = \mathbf{V}_R \mathbf{V}_R^T \mathbf{x} \quad (20)$$

The information lost by the dimensional compression, that is, errors, is written as

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}} = (\mathbf{I} - \mathbf{V}_R \mathbf{V}_R^T) \mathbf{x}. \quad (21)$$

Using the errors, the Q statistic is defined as

$$Q = e^T e \quad (22)$$

$$= x^T (I - V_R V_R^T) x. \quad (23)$$

The Q statistic is the distance between the sample and the subspace spanned by principal components. In other words, the Q statistic is a measure of dissimilarity between the sample and the modeling data from the viewpoint of the correlation among variables. Therefore, the class that minimizes the Q statistic of x should be selected as the class of x . The classifier h is described as

$$h(x) = \arg \min_j Q_j \quad (24)$$

$$Q_j = x^T (I - V_R^{(j)} V_R^{(j)T}) x \quad (25)$$

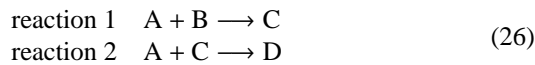
where $V_R^{(j)}$ is the loading matrix derived from the matrix $X^{(j)}$ whose rows are samples belonging to class K_j .

5. Case study

In this section, the proposed NC-spectral clustering is compared with the k -means method through their application to operation data of parallelized batch processes. In addition, soft-sensors for product composition are designed on the basis of their clustering results. The detailed batch process model used in this case study is described in Appendix A.

5.1. Problem setting

A schematic diagram of the batch reactor is shown in Fig. 7. In this process, a well-mixed, liquid-phase reaction system is considered, in which two reactions take place:



The component C is the desired product while D is the byproduct, and the objective is to achieve a good conversion of product C. The initial reactor temperature is 20 °C and the initial amount of the raw materials A and B changes as the random numbers following $N(20, 0.1)$.

Since reaction 1 proceeds at 90 °C or higher, the reactor temperature should be raised as soon as possible after operation starts. As reaction 1 proceeds, the reactor temperature increases due to reaction heat. However, reaction 2 proceeds at the reactor temperature exceeding 100 °C and product C is converted to byproduct D. Therefore, a rise in the reaction temperature has to be controlled after it reaches around 90 °C. In this process,

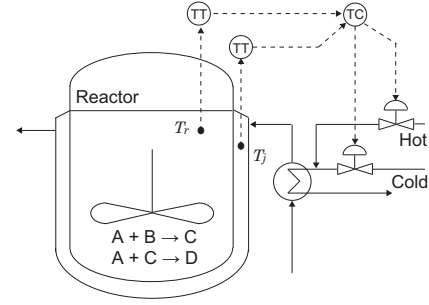


Figure 7: Schematic diagram of the batch reactor with control systems

the reactor temperature and the jacket temperature are controlled through the multivariable control system.

The jacket temperature T_j and the reactor temperature T_r are measured every one minute and the termination time of operation is 120 minutes. The operation data of the batch process is shown as Fig. 8.

In this case study, five reactors R_1, \dots, R_5 are operated in parallel. In addition, their heat transfer coefficients are unchanged during the batch operation, and they change every batch operation as the random number. The heat transfer coefficients U_i ($i = 1, \dots, 5$) are

$$U_i = \begin{cases} U(40.60, 40.62), & i = 1, 2 \\ U(40.57, 40.59), & i = 3, 4 \\ U(40.54, 40.56), & i = 5 \end{cases} \quad (27)$$

where $U(a, b)$ denotes the uniform random numbers on the closed interval $[a, b]$. Therefore, there are three types of the correlation among variables although there are five reactors. In this case study, the operation data of 20 batches of each reactor are stored in the database.

The objective of this case study is to construct soft-sensors that can estimate the amount of the product C accurately.

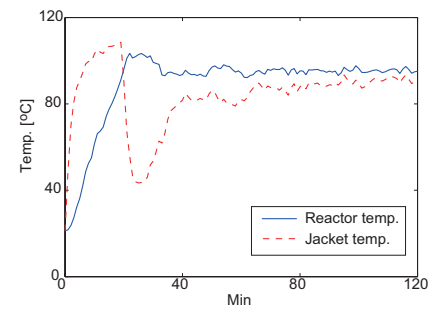


Figure 8: Batch process operation data

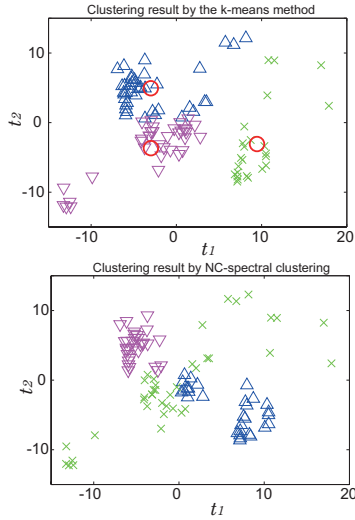


Figure 9: Classification results: the k -means method (left) and NC-spectral clustering (right)

5.2. Clustering

Before constructing soft-sensors, the operation data of all 100 batches stored in the database are clustered into three classes using the k -means method and the proposed NC-spectral clustering. As preprocessing of the operation data, its dimension is reduced by Multiway PCA [20, 21] and the number of the retained principal components is two. Therefore, the input variables of these clustering methods are the scores t_1 and t_2 . The parameter of NC-spectral clustering is $\gamma = 0.99$.

The clustering results of the k -means method and NC-spectral clustering are shown in Fig. 9. In this figure, the sample distribution on the t_1 - t_2 plane is shown. In the result of the k -means method, the center of each class \mathbf{c}_j ($j = 1, 2, 3$) is designated by a circle, and samples are certainly classified based on the distance. On the other hand, in the case of the proposed method, samples are classified regardless of the distance.

5.3. Soft-sensor design

A soft-sensor is constructed to estimate the amount of product M_C [kmol] at the end of the batch. Three models are built for each of three classes clustered by NC-spectral clustering in Sec. 5.2. The input variables of soft-sensors are the time series of T_r and T_j , and PLS is used for model construction.

In addition, another soft-sensor is designed on the basis of the clustering result of the k -means method. In this case, the distance from the class center \mathbf{c}_j is used for sample discrimination when an output is requested.

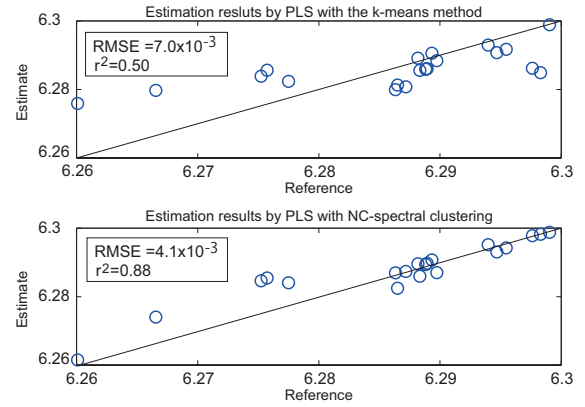


Figure 10: Estimation results by PLS with the k -means method (top) and NC-spectral clustering (bottom)

That is, the classifier g is described as

$$g(\mathbf{x}) = \arg \min_j \|\mathbf{x} - \mathbf{c}_j\|_2. \quad (28)$$

In this case study, M_C of the new reactor R_6 is estimated for performance validation. The heat transfer coefficient of R_6 changes as the random number following $U(40.60, 40.62)$, which is the same as that of R_1 and R_2 . The number of the validation batches is 20.

The estimation results of soft-sensors are shown in Fig. 10. In these figures, the horizontal line and the vertical line express the measurement and the estimates, respectively. RMSE is the root-mean-squared error and r^2 denotes the determination coefficient between measurements and estimates. These results clearly show that the proposed method can achieve higher estimation performance than the k -means method-based soft-sensor and RMSE is improved by 43%.

6. Conclusion

A new clustering method is proposed by integrating the NC method and spectral clustering. The proposed NC-spectral clustering can accurately discriminate the correlation among variables. In addition, a new soft-sensor design method based on NC-spectral clustering is proposed. Since NC-spectral clustering can discriminate the individuality of production devices, the proposed soft-sensor design method can improve the estimation performance. The usefulness of the proposed methods is demonstrated through numerical examples and a case study of parallelized batch processes.

References

- [1] T. Mejdell and S. Skogestad, Estimation of Distillation Compositions from Multiple Temperature Measurements Using Partial-

- Least-Squares Regression, Ind. Eng. Chem. Res. 30 (1991) 2543-2555.
- [2] M. Kano, K. Miyazaki, S. Hasebe and I. Hashimoto, Inferential Control System of Distillation Compositions Using Dynamic Partial Least Squares Regression, J. Process Control 10 (2000) 157-166.
- [3] H. Kamohara, A. Takinami, M. Takeda, M. Kano, S. Hasebe and I. Hashimoto, Product Quality Estimation and Operating Condition Monitoring for Industrial Ethylene Fractionators, J. Chem. Eng. Jpn. 37 (2004) 422-428.
- [4] H. Kaneko, M. Arakawa and K. Funatsu, Development of a new soft sensor method using independent component analysis and partial least squares, AIChE J. 55 (2009) 87-98.
- [5] V. Radhakrishnan and A. Moham, Neural networks for the identification and control of blast furnace hot metal quality, J. Process Control 10 (2000) 509-524.
- [6] R. Amirthalingam and J. Lee, Subspace Identification Based Inferential Control Applied to a Continuous Pulp Digester, J. Process Control 9 (1999) 397-406.
- [7] M. Kano, S. Lee, and S. Hasebe, Two-Stage Subspace Identification for Softsensor Design and Disturbance Estimation, J. Process Control 19 (2008) 179-186.
- [8] M. Kano and M. Ogawa, The State of the Art in Advanced Chemical Process Control in Japan, Proc. of ADCHEM, CD-ROM, Istanbul, Turkey 2009.
- [9] S. J. Qin, Recursive PLS Algorithms for Adaptive Data Modeling, Comput. Chem. Eng. 22 (1998) 503-514.
- [10] G. Bontempi, M. Birattari and H. Bersini, Lazy Learning for Local Modeling and Control Design, Int. J. Control 72 (1999) 643-658.
- [11] C. G. Atkeson, A. W. Moore and S. Schaal, Locally Weighted Learning, Artif. Intell. Rev. 11 (1997) 11-73.
- [12] C. Cheng, and M. S. Chiu, A New Data-Based Methodology for Nonlinear Process Modeling, Chem. Eng. Sci. 59, (2004) 2801-2810.
- [13] K. Fujiwara, M. Kano and S. Hasebe, Soft-Sensor Development using Correlation-Based Just-In-Time modeling, AIChE J. 55, (2009) 1754-1765.
- [14] T. Kohonen, Self-organizing maps, 3rd edn, Springer, New York, 2001.
- [15] K. Fujiwara, M. Kano and S. Hasebe, Correlation-Based Pattern Recognition Algorithm and Its Application to Adaptive Software Sensing, Proc. of ADCHEM, CD-ROM, Istanbul, Turkey 2009.
- [16] C. Ding, X. He, H. Zha, M. Gu, and H. D. Simon, A Min-Max Cut Algorithm for Graph Partitioning and Data Clustering, Proc. of IEEE Int'l Conf. Data Min. San Jose 2001.
- [17] A. N. Ng, M. I. Jordan and Y. Weiss, On Spectral Clustering: Analysis and an algorithm, Proc. of Neural Inf. Process. Syst. Vancouver 2001.
- [18] M. Fiedler, Algebraic connectivity of graphs, Czech. Math. J. 23, (1973) 298-305.
- [19] J. E. Jackson, and G. S. Mudholkar, Control Procedures for Residuals Associated with Principal Component Analysis, Technometrics, 21, (1979) 341-349.
- [20] P. Nomikos and J. F. MacGregor, Monitoring Batch Processes Using Multiway Principal Component Analysis, AIChE J. 40 (1994) 1361-1375.
- [21] P. Nomikos and J. F. MacGregor, Multiway Partial Least Squares in Monitoring Batch Processes, Chemometr. Intell. Lab. Syst. 30 (1995) 97-109.
- [22] B. J. Cott and S. Macchietto, Temperature Control of Exothermic Batch Reactors Using Generic Mode Control, Ind. Eng. Chem. Res. 28, (1989) 1177-1184.

Appendix A. Batch process model

The batch process model used in the case study is described in [22]. The balance equations around reactor and jacket are as follows:

$$\frac{dM_A}{dt} = -R_1 - R_2 \quad (\text{A.1})$$

$$\frac{dM_B}{dt} = -R_1 \quad (\text{A.2})$$

$$\frac{dM_C}{dt} = R_1 - R_2 \quad (\text{A.3})$$

$$\frac{dM_D}{dt} = R_2 \quad (\text{A.4})$$

$$R_1 = k_1 M_A M_B \quad (\text{A.5})$$

$$R_2 = k_2 M_A M_C \quad (\text{A.6})$$

$$k_1 = A_1 \exp(-E_1/RT'_r) \quad (\text{A.7})$$

$$k_2 = A_2 \exp(-E_2/RT'_r) \quad (\text{A.8})$$

$$W = M_{wA}M_A + M_{wB}M_B + M_{wC}M_C + M_{wD}M_D \quad (\text{A.9})$$

$$M_r = M_A + M_B + M_C + M_D \quad (\text{A.10})$$

$$C_{PR} = (C_{pA}M_A + C_{pB}M_B + C_{pC}M_C + C_{pD}M_D)/M_r \quad (\text{A.11})$$

$$V = W/\rho \quad (\text{A.12})$$

$$Ar = V/D \quad (\text{A.13})$$

$$Q_j = UA(T'_j - T'_r) \quad (\text{A.14})$$

$$Q_r = -\Delta H_1 R_1 - \Delta H_2 R_2 \quad (\text{A.15})$$

$$\frac{T'_r}{dt} = \frac{Q_r + Q_j}{M_r C_{pr}} \quad (\text{A.16})$$

$$\frac{T'_j}{dt} = \frac{F_j \rho_j C_{pj} (T_j^{sp} - T'_j) - Q_j}{V_j \rho_j C_{pj}} \quad (\text{A.17})$$

$$T_r = T'_r + a^{(k)} - 0.866a^{(k-1)} \quad (\text{A.18})$$

$$T_j = T'_j + a^{(k)} - 0.866a^{(k-1)} \quad (\text{A.19})$$

where $a^{(k)} \sim N(0, 0.1)$ and k denotes the measurement step number. The dynamic of control valves for jacket temperature is modeled as Eq.(A.20), and PI controllers are used.

$$P_V(s) = \frac{K}{\tau s + 1} \quad (\text{A.20})$$

where $K = 1/16$ and $\tau = 2\text{min}$. The model parameters and the initial condition are given in Tables A.2 and A.3.

Table A.2: Model parameters for the batch process

$M_{wA} = 30 \text{ kg/kmol}$	$C_{pA} = 75.31 \text{ kJ/kmol} \cdot \text{K}$
$M_{wB} = 100 \text{ kg/kmol}$	$C_{pB} = 167.31 \text{ kJ/kmol} \cdot \text{K}$
$M_{wC} = 130 \text{ kg/kmol}$	$C_{pC} = 217.57 \text{ kJ/kmol} \cdot \text{K}$
$M_{wD} = 160 \text{ kg/kmol}$	$C_{pD} = 334.73 \text{ kJ/kmol} \cdot \text{K}$
$k_1^0 = 1.20 \times 10^9 [-]$	$\Delta H_1 = -41,840 \text{ kJ/kmol}$
$k_2^0 = 7.88 \times 10^{16} [-]$	$\Delta H_2 = -25,105 \text{ kJ/kmol}$
$E_1/R = 1.00 \times 10^4 [-]$	$\rho = 1.00 \times 10^3 \text{ kg/m}^3$
$E_2/R = 1.70 \times 10^4 [-]$	$\rho_j = 1.00 \times 10^3 \text{ kg/m}^3$
$r = 0.50 \text{ m}$	$C_{pj} = 1.88 \text{ kJ/kg} \cdot \text{K}$
$F_j = 0.38 \text{ m}^3/\text{min}$	$V_j = 0.69 \text{ m}^3$
$U = 40.842 \text{ kJ/min} \cdot \text{m}^2 \cdot \text{K}$	

Table A.3: Initial operating condition for the batch process

$M_{A,0} = 12 \text{ kmol}$	$M_{B,0} = 12 \text{ kmol}$
$M_{C,0} = 0 \text{ kmol}$	$M_{D,0} = 0 \text{ kmol}$
$T_{r,0} = 293.15 \text{ K}$	$T_{j,0} = 293.15 \text{ K}$